Economic Impacts of Wind Covariance Estimation on Power Grid Operations

Cosmin G. Petra, Victor M. Zavala,
Mihai Anitescu, and Elias D. Nino-Ruiz

***Abstract*—We study the impact of capturing spatiotemporal correlations between multiple supply points on economic dispatch procedures. Using a simple dispatch model we first show analytically that over/underestimation of correlation leads to positive and negative biases of dispatch cost. A rigorous, large-scale computational study for the State of Illinois transmission grid with real topology and physical constraints reveals similar conclusions. For this study, we use the Rao-Blackwell-Ledoit-Wolf estimator to approximate the wind covariance matrix from a small number of wind samples generated with the numerical weather prediction model WRF and we use the covariance information to generate a large number of wind scenarios. The resulting stochastic dispatch problems are solved using the interior-point solver PIPS-IPM on the BlueGene/Q (Mira) supercomputer at Argonne National Laboratory. Our results indicate that strong and persistent biases result from neglecting correlation information and points to the need to design a market that coordinates forecasts and uncertainty information.**

***Index Terms*—covariance, correlation, spatiotemporal, estimation, uncertainty, wind power, dispatch**

NOMENCLATURE

Wind Forecast	
γ	Ensemble inflation factor
Dispatch	
$\mathcal{N}$	set of nodes
$\mathcal{G}$	set of suppliers
$\mathcal{U}$	feasible set for flows
$\mathcal{C}_i$	feasible set for supplier i
Ω	set of scenarios
x_i	forward dispatch quantity for supplier i
$X_i(\omega)$	spot dispatch quantity for supplier i in scenario ω
f	forward flow vector
$F(\omega)$	spot flow vector in scenario ω
p_i	cost for supplier i
p_i^+	additional supply cost for supplier i
p_i^-	buyback cost for supplier i
d_n	demand at node n
$\tau_n(\cdot)$	injection function for node n
M_i	supply limit for deterministic supplier i
$W_i(\omega)$	supply limit for stochastic supplier i in scenario ω

I. INTRODUCTION

Achieving efficient grid operations under large-scale wind and solar power is a challenge because these supply sources follow complex spatiotemporal patterns (see Figure fig:corrwind) that extend over wide geographical regions (e.g., tens to hundreds of kilometers) and long periods of time (i.e., hours to days). Because of this, reserve allocation procedures can be ineffective and more adaptive and systematic approaches based on stochastic and robust optimization techniques are needed.

Stochastic and robust optimization techniques rely on uncertainty characterizations. Correlation (or covariance) information, in particular, is key because this guides forecast aggregation/disaggregation procedures and because it is needed to characterize risk in dispatch cost and revenues of market players. If the supply of a wind farm in a region is uncorrelated from that in another region, these can be forecasted independently without affecting dispatch cost. However, when correlations exist, one would expect that using independent forecasts introduces errors in the uncertainty characterization, and will bias dispatch cost and will shift incentives of the players (wind power suppliers, suppliers, and consumers). This was hypothesized in [10] (see Section 5), however, before the present study, there was not any evidence that this is case or how large the cost bias is.

Estimating long-range correlations is challenging from a market implementation point of view and from a computational point of view. From a market implementation stand-point, determining long-range correlations is a challenge because wind farm owners might not be willing to share their forecasting procedures and their site information with other markets players and the ISO. Consequently, they might prefer to construct their own forecasts and uncertainty levels by possibly neglecting correlations with other sites. Wind farm owners might also need to neglect long-range correlations altogether simply because of computational limitations faced by their forecasting vendors. Computational challenges arise because properly resolving the space-time resolution of numerical weather prediction (NWP) systems requires significant computational power [3]. NWP systems are extremely computationally intensive and few computing sites exist in the world that can obtain forecasts that accurately capture both short-range wind farm conditions and long-range behavior. In other words, there are limits on the resolution of uncertainty characterizations achievable and this leads to ambiguity. Note also that keeping forecasting information confidential provides a mechanism for manipulation (e.g., a supplier overestimates its uncertainty).

Properly designing decentralized markets that factor in uncertainty in weather-driven supply is necessary but this requires significantly more complex information exchange mechanisms between the ISO and market players compared to existing deterministic settings [10], [17]. To design such information exchange mechanisms, it is necessary to understand the effects that certain pieces of information have on market performance.

C. G. Petra, V. M. Zavala and M. Anitescu are with the Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. E-mail: {petra, vzavala, anitescu}@mcs.anl.gov.

E. D. Nino-Ruiz is with the Department of Computer Science, Virginia Tech, Blacksburg, VA 24061. E-mail: enino@vt.edu

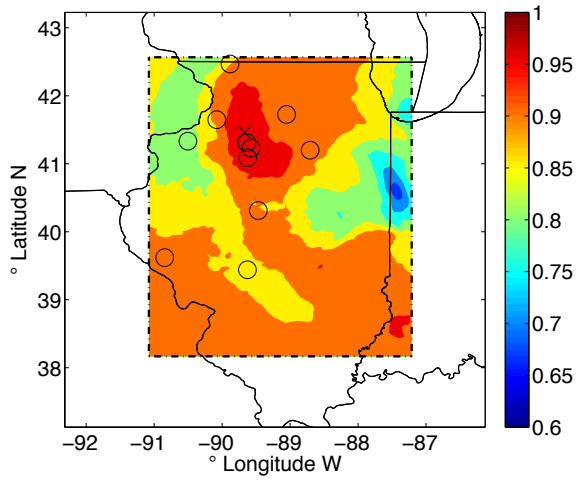

Fig. 1. Spatial correlation for wind speed in State of Illinois.

In this work, we study the particular effect of long-range correlation information on dispatch cost. We first present an analytic example to prove that correlation between suppliers output can positively or negatively bias dispatch cost (depending on the correlation direction). Second, we perform a detailed computational study using a stochastic economic dispatch setting in the Illinois transmission system. We use validated wind speed ensemble forecasts obtained with the NWP system WRF that are propagated through typical wind power curves to obtain wind power ensembles. We use the Rao-Blackwell-Ledoit-Wolf estimator to generate scenarios from the limited number of WRF samples available and demonstrate the efficacy of this estimator. Our computation study reveals that: i) dispatch cost biases that would scale up (if similar correlation patterns across the year hold) to order of hundreds of millions of dollars a year can be introduced by ignoring long-range correlation and ii) that confidence levels of dispatch cost are different from the actual ones when correlations are neglected. In our study, the confidence intervals when ignoring correlations were wider thus requiring many more scenarios to close the gap compared to the correlation-based approach. Our study thus indicates that, as hypothesized in [10], centralized forecasts that can properly account for correlations are superior to localized ones when used for constructing wind power bids by suppliers in markets with signficant wind power penetration.

II. A Motivating Analytical Example

Consider a single-node system with three suppliers and one demand. The first two suppliers, G_1 and G_2, are considered to have uncertain power output. Assume that the power outputs for these suppliers follow Gaussian distributions, $\mathcal{N}(w_1, \sigma_1)$ and $\mathcal{N}(w_2, \sigma_2)$, and define $\rho \in [-1, 1]$ as the correlation coefficient. Assume also that both supply power at cost p_w. The third supplier G_3 is assumed to be deterministic; this supplies power at cost p_{th} with $p_{th} > p_w$ and has infinite capacity. The demand quantity is defined as d and we assume this to be deterministic and inelastic.

By construction, it can be deduced that as much cheaper power as possible should be produced. In case this does not satisfy all demand, then G_3 will be dispatched to fulfill the remaining demand. Consequently, the negative dispatch cost is

$$c_d = \mathbb{E}\left[p_w \min(X_1 + X_2, d) + p_{th} \max(d - X_1 - X_2, 0)\right]. \tag{II.1}$$

To show the dependence $c_d = c_d(\rho)$ we write (II.1) as follows:

$$\begin{aligned} c_d &= \mathbb{E}\left[p_w d + p_w \min(X_1 + X_2 - d, 0)\right. \\ &\qquad \left. + p_{th} \max(d - X_1 - X_2, 0)\right] \\ &= p_w d + \mathbb{E}\left[-p_w \max(d - X_1 - X_2, 0)\right. \\ &\qquad \left. + p_{th} \max(d - X_1 - X_2, 0)\right] \\ &= p_w d + \mathbb{E}\left[(p_{th} - p_w) \max(d - X_1 - X_2, 0)\right] \\ &= p_w d + (p_{th} - p_w)\mathbb{E}\left[d - (X_1 + X_2) \,|\, X_1 + X_2 \le d\right]. \end{aligned} \tag{II.2}$$

Here, $\mathbb{E}[X|Y]$ denotes the expectation of X conditional on event Y. Furthermore, since the random variable $X = X_1 + X_2$ is normally distributed, $X \sim \mathcal{N}(\mu, \sigma)$, where

$$\begin{aligned} \sigma &= \sigma(\rho) \\ &= \sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2} \text{ and } \mu = w_1 + w_2, \end{aligned} \tag{II.3}$$

c_d can be expressed

$$\begin{aligned} c_d &= p_w d + (p_{th} - p_w)\mathbb{E}\left[d - X | X \le d\right] \\ &= p_w d + (p_{th} - p_w) \cdot d \cdot \Phi(d, \sigma) - (p_{th} - p_w)\mathbb{E}\left[X | X \le d\right], \end{aligned} \tag{II.4}$$

where Φ is the cumulative density function of X,

$$\Phi(x, \sigma) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x - \mu}{\sqrt{2}\sigma}\right)\right]. \tag{II.5}$$

Also let us denote the probability density function of Y by $\phi(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. It can be shown (see Lemma 1 from the Appendix) that

$$\mathbb{E}\left[X | X \le d\right] = -\sigma^2 \phi(d, \sigma) + \mu\, \Phi(d, \sigma). \tag{II.6}$$

Combining this equation with (II.4), the dispatch cost is

$$c_d(\sigma) = p_w d + (p_{th} - p_w)\left((d - \mu)\Phi(d, \sigma) + \sigma^2 \phi(d, \sigma)\right). \tag{II.7}$$

The dependence of the dispatch cost c_d on the correlation parameter ρ becomes clear an is illustrated in Figure 7. We also have the following result:

Proposition 1: The dispatch cost $c_d(\rho)$ is a strictly increasing function of the correlation parameter ρ.
Proof: See Appendix A.

This result shows that there can be a positive or negative bias (depending on the correlation direction) between dispatch cost values computed from different correlation structures (different uncertainty characterizations). This can be interpreted as a dispatch cost bias introduced by an error in the correlation coefficient resulting from computational or market implementation limitations.

With increasing ρ, the probability distribution of the total available wind $X = X_1 + X_2$ becomes "taller" (see (II.3)), and, as a consequence, it is less likely to satisfy demand using the cheap uncertain supply and more probable to dispatch the

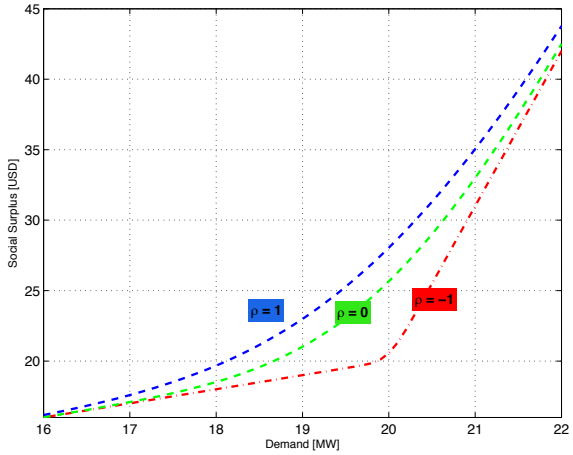

Fig. 2. Dispatch cost c_d as a function of demand d for negative correlation ($\rho = -1$), no correlation ($\rho = 0$) and positive correlation ($\rho = 1$).

more expensive supply and this results in a higher cost. For instance, if two suppliers are fully positively correlated ($\rho = 1$), a low output of one will result in a low output of the other. In the extreme case of full negative correlation ($\rho = -1$) cheap uncertain supply will most likely be used because if the output of of one wind farm drops then this implies that the output of the other one increases so there is always at least one active wind farm. From Figure 7 note also that, at low and high demands, correlation does not affect dispatch cost.

If analyzed from a purely dispatch cost perspective, one can argue that an inaccurate forecasting system that overestimates and underestimates the correlations may lead to the same expected dispatch cost on the long run compared to that obtained with an ideal forecasting system. While we do not refute this possibility, we point out that inaccurate forecasting also results in gaps in prices and revenues and, for this reason, it results in *market inefficiencies*. This will bias incentives in unpredictable ways towards certain players that might not average out in the long run.

Errors in correlation between uncertain suppliers can occur for a variety of reasons. For example, the owners of the wind farms might submit their own forecasts and scenarios and the ISO will treat them independently when dispatching. Such a situation is discussed in [11]. Alternatively, correlations or the covariance matrix may be poorly estimated in the NWP systems due to the prohibitively expensive computational cost of accurate estimators. Hence, we advocate the use of ISO-centric weather forecasting system of atmospheric conditions while mapping to wind power forecasts (that capture local and wind turbine effects) can be done internally by wind farm owners in order to prevent disclosure of information. The numerical simulations of the next section indicate that more than $\$10,000$ savings per dispatch period can be obtained for full-scale power grids, such as the one of the State of Illinois. As an estimate of the savings over one year, it can reach a hundred millions dollars, assuming the same amount of savings per dispatch period as the one revealed by our simulations.

III. Detailed Computational Study

In this section, we present a detailed computational framework to analyze the effects of correlation on dispatch cost and prices. The different elements of the framework include i) a stochastic dispatch model, ii) numerical weather prediction, and iii) covariance estimation and scenario generation. Our setting includes realistic data for the Illinois transmission system and we assume a wind adoption scenario of 17%.

A. Stochastic Dispatch Model

In our analysis we use the stochastic dispatch model from [11] because it captures forward and spot market components in a realistic manner.

$$\min_{x_i, X_i(\omega)} \sum_{i \in \mathcal{G}} \left(p_i x_i + \mathbb{E}_\omega \left[p_i^+ (X_i(\omega) - x_i)_+ - p_i^- (X_i(\omega) - x_i)_- \right] \right) \tag{III.8a}$$

s.t.

$$\tau_n(f) + \sum_{i \in \mathcal{G}(n)} x_i = d_n, \quad n \in \mathcal{N} \tag{III.8b}$$

$$\tau_n(F(\omega)) - \tau_n(f) + \sum_{i \in \mathcal{G}(n)} (X_i(\omega) - x_i) = 0, \quad n \in \mathcal{N}, \omega \in \Omega \tag{III.8c}$$

$$f, F(\omega) \in \mathcal{U}, \omega \in \Omega \tag{III.8d}$$

$$(x_i, X_i(\omega)) \in \mathcal{C}_i(\omega), \quad i \in \mathcal{G}, \omega \in \Omega \tag{III.8e}$$

Here, N denotes the set of nodes (buses) and L the set of transmission lines that form the power grid. The set of all suppliers is denoted by $\mathcal{G}$. Subsets $\mathcal{G}(n)$ denote the set of players connected to node n. The forward dispatched quantities for players are x_i and the spot quantities under scenario ω are $X_i(\omega)$. The forward power flow through line $\ell \in L$ is denoted by f_ℓ and f denotes the vector of all line flows. Similarly, $F(\omega)$ denotes the vector of line flows $F_\ell(\omega)$ for each scenario ω. The demand is assumed to be deterministic and inelastic and represented by d_n, $n \in N$.

The scenarios ω characterize the randomness in the model due to unpredictable capacities and are mathematically are expressed as random vectors defined on some probability space $(\Omega, \mathcal{F}, P)$. The expectation $\mathbb{E}_\omega$ is taken with respect to the measure P. In practice, one considers a finite approximation of Ω obtained through sampling.

The objective consists of minimizing the forward dispatch cost $\sum_{i \in \mathcal{G}} p_i x_i$ plus the expected adjustment or recourse dispatch cost $\sum_{i \in \mathcal{G}} \mathbb{E}_\omega \left[p_i^+ (X_i(\omega) - x_i)_+ - p_i^- (X_i(\omega) - x_i)_- \right]$ specific to individual scenarios. Here $[y]_+ = \max\{y, 0\}$ and $[y]_- = \max\{-y, 0\}$. The coefficients p_i denote the bid price and p_i^+ and p_i^- are price bids for real-time corrections of the generators. A supplier i asks $p_i^+ > p_i$ to sell additional power, or asks $p_i^- < p_i$ to buy power from the system (e.g., reduces output). In our model we have used $p_i^- = 0.8 p_i$ and $p_i^+ = 1.2 p_i$.

The decision variables are the power dispatch of each generator, x_i, $X_i(\omega)$ and the power flows f_l and $F(\omega)$. The first-stage dispatches x_i are "ahead" decisions that accounts for randomness; the second-stage re-dispatches $X_i(\omega)$ represent

"real time" decisions that are appropriate corrections once an individual realization ω of the randomness is observed.

Function $\tau_n(\cdot)$ is a mapping of the flow vector to the node n. We denote by $\nu_1(n)$ the inflow lines into node $n \in N$ and by $\nu_0(n)$ the outflow lines. Equation (III.8b) describes the power flow through a node $n \in N$ which is the sum of power $\tau_n(f) = \sum_{l\in\nu_1(n)} f_l - \sum_{l\in\nu_0(l)} f_l$ imported via the transmission lines to node n and power $\sum_{i\in\mathcal{G}(n)} x_i$ produced at node n. Equation (III.8c) is the second-stage correspondent of (III.8b), enforcing power flow balance at each node for each scenario ω. It is shown in [11] that the multipliers associated with this "residual" formulation (not with the simpler equivalent form, $\tau_n(F(\omega)) + \sum_{i\in\mathcal{G}(n)} X_i(\omega) = 0$) gives the clearing prices to be used in the spot (or real-time) market. Equation (III.8d) represents maximum flow constraints on individual lines, with $\mathcal{U}$ being usually a polyhedron.

Equation (III.8e) expresses constraints on supply that come from technological limits of the generators (such as maximum/minimum capacity, limited ramp-up/down power on short notice) and intermittent availability of energy of some generators. In our model, $\mathcal{C}_i$ is deterministic for thermal generators (natural gas, coal, heater oil and nuclear) and given by $\mathcal{C}_i = \{(x_i, X_i) : x_i, X_i \in [m_i, M_i], |x_i - X_i| \leq r_i\}$, expressing capacity and ramp-up constraints. The uncertain output is modeled by $\mathcal{C}_i(\omega) = \{(x_i, X_i(\omega)) : x_i \in [0, M_i], X_i(\omega) \in [0, W_i(\omega)]\}$, showing that the first-stage dispatch x_i can be allowed to reach maximum installed capacity and the second-stage dispatch $X_i(\omega)$ can only be allowed to reach maximum power generated under scenario ω. We consider ramp constraints with only one time step.

Our model was set up for the State of Illinois power grid which comprises 2522 lines, 1908 buses, 870 load buses, 225 generators, of which 32 are wind farms. In order to obtain a large wind power installed capacity (approximately 17%), we needed to create synthetic wind farms in addition to the existing ones. The synthetic farms were chosen to replace existing coal or gas generators. This was done specifically to avoid possible network congestions that would limit the amount of real wind adoption. In addition, we replaced only thermal generators that were mirrored by other (usually identically) generators to ensure that enough thermal generation is available to satisfy demand when low-wind scenarios occur. The cost per MW for the wind farm is 5$/MW, the lowest across all generators.

B. Wind Scenario Generation

Wind direction and speed samples required for our study are obtained from WRF. The WRF model [13] is a state-of-the-art numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. WRF is the result of a multi agency and university effort to build a highly parallelizable code that can run across scales ranging from large-eddy to global simulations. WRF has a comprehensive description of the atmospheric physics that includes cloud parameterization, land-surface models, atmosphere-ocean coupling, and broad radiation models.

We set up a computational nested domain structure for WRF including a high-resolution sector that covers the State of Illinois and two additional domains of larger coverage but lower resolution that provide boundary conditions for the nested domains [2]. This is illustrated in Figure 4. The initial conditions from the assimilated state (also known as reanalyzed state) of the atmosphere are randomly perturbed and propagated in time through the nonlinear NWP model to obtain a set of ensembles that describe possible trajectories of the atmospheric conditions. The computational cost of this procedure is significant. Computing a single ensemble for Illinois over a 24-hour time window takes 6 hours of wall clock time running on 32 processors. Because of these limitations, we computed only 30 ensembles for June 4, 2006 in [2]. The ensembles have been validated using weather station observations obtained from the National Climatic Data Center (NCDC). The ensembles for six different wind farm locations are shown in Figure .

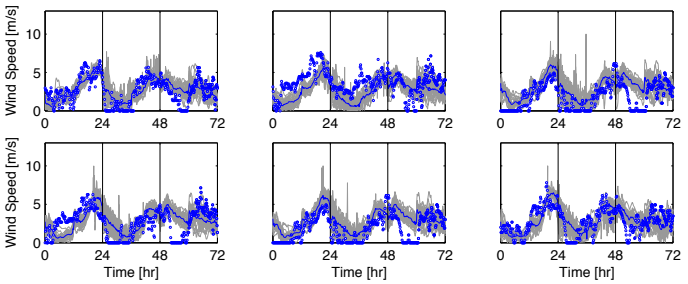

Fig. 3. Wind-speed realizations for 6 wind farm locations in Illinois and observations (dots) at nearest meteorological stations. Vertical lines represent beginning of day (12:00 AM).

Each ensemble provides the components of the wind velocity which are transformed to wind speed. This gives a 3-D field in geographical coordinates (latitude, longitude, and elevation) evolving over time where the field points match the discretization mesh in the inner domain. The wind farm locations, however, do not match the discretization mesh. In addition, the typical hub height used for wind farms (80 meters) may also not match the WRF vertical layers. To remedy this issue we use linear and bilinear interpolation to compute wind speeds at the farm locations from the WRF ensembles, therefore obtaining a set of 30 ensembles for the speeds at the desired 3-D coordinates. We then compute the sample mean $\bar{\mathbf{x}}$, sample covariance $\mathbf{S}$ and the Rao-Blackwell-Ledoit-Wolf (RBLW) estimator $\widehat{\mathbf{\Sigma}}_{RBLW}$ as described in the next section. Using this approach, we can compute many wind speed scenarios by sampling a multivariate Gaussian with mean $\bar{\mathbf{x}}$ and estimated covariance matrix $\widehat{\mathbf{\Sigma}}_{RBLW}$.

Our computational framework is not restricted to the use of WRF and RBLW covariance estimators and can accommodate scenarios generated with different scenario generation techniques for wind speed, assuming that these can handle the large spatial coverage present in our computational study. Among alternative scenario generation techniques we mention copula transformation [15], [16], probabilistic models [14] and time-recursive estimation of covariance matrix [9].

C. Covariance Estimation Methods and Validation

The simplest estimator for the covariance using the WRF samples would be the sample covariance matrix. In our setting, however, this matrix is rank-deficient and therefore noninvertible

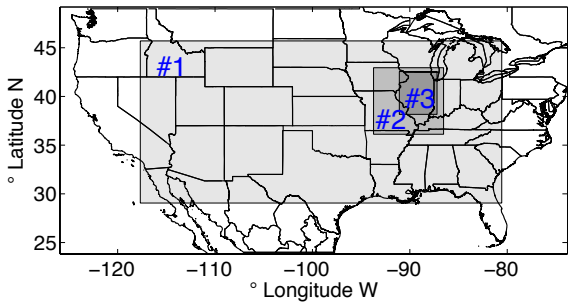

Fig. 4. WRF simulation domain with resolutions of #1 - 32 km^2, #2 - 6 km^2, and #3 - 2 km^2.

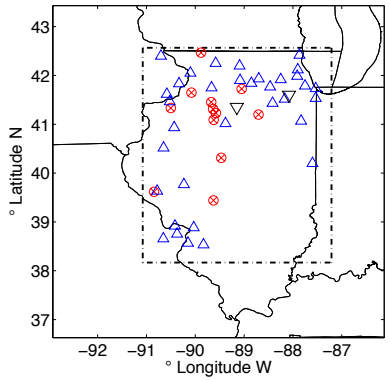

Fig. 5. Locations of wind farms (blue triangles) and meteorological stations (red dots).

since the number of WRF samples is smaller than the number of the wind farms. Because of this, it is not possible to factorize it and sample from it. In addition, the sample covariance estimator can be unreliable when few samples are available. Consequently, we investigate the use of alternative estimators and we compare their performance.

Mathematically, the problem of covariance estimation can be expressed as the problem of estimating the matrix $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ of a random vector $\mathbf{y} \in \mathbb{R}^p$ based on a set of realizations or samples $\{\mathbf{x}_i\}_{i=1}^n$, for the case when n is smaller than p. Let $\bar{\mathbf{x}} = \frac{1}{n}\cdot\sum_{i=1}^n \mathbf{x}_i \in \mathbb{R}^p$ denote the sample mean and $\mathbf{S} = \frac{1}{n}\cdot\mathbf{\Delta X}\cdot\mathbf{\Delta X}^T \in \mathbb{R}^{p\times p}$ denote the sample covariance matrix, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n] \in \mathbb{R}^{p\times n}$ and $\mathbf{\Delta X} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}] \in \mathbb{R}^{p\times n}$.

Covariance estimators for the case $p > n$ make use of regularizations of the sample covariance matrix $\mathbf{S}$ to overcome the rank-deficiency. One of the most common regularizations is the perturbation of the covariance matrix by a multiple of the identity, which leads to shrinkage estimators of the type

$$\widehat{\mathbf{\Sigma}} = \alpha\cdot\mathbf{I} + \beta\cdot\mathbf{S}, \tag{III.9}$$

where $\mathbf{I} \in \mathbb{R}^{p\times p}$ is the identity matrix and α and β are regularization parameters or weights chosen to minimize the estimation error $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$, where $\|\cdot\|$ usually refers to the Frobenius norm. Since $\mathbf{\Sigma}$ is unobservable, different techniques are used to compute statistical estimates for α and β.

Ledoit and Wolf [4] propose the estimator $\widehat{\mathbf{\Sigma}}_{LW}$ given by

$$\widehat{\mathbf{\Sigma}}_{LW} = \frac{b^2}{d^2}\cdot m\cdot\mathbf{I} + \frac{a^2}{d^2}\cdot\mathbf{S},$$

where,

$$m = \mathrm{tr}\left(\mathbf{S}\cdot\mathbf{I}\right)/p, \quad d^2 = \mathrm{tr}\left(\left[\mathbf{S} - m\cdot\mathbf{I}\right]\cdot\left[\mathbf{S} - m\cdot\mathbf{I}\right]^{\mathrm{t}}\right)/p, \tag{III.10}$$

$$b^2 = \min\left(n^{-1}\cdot\sum_{k=1}^{n}\frac{\mathrm{tr}\left(\left[\mathbf{x}_k\cdot\mathbf{x}_k^{\mathrm{t}} - \mathbf{S}\right]\cdot\left[\mathbf{x}_k\cdot\mathbf{x}_k^{\mathrm{t}} - \mathbf{S}\right]^{\mathrm{t}}\right)}{p}, d^2\right), \tag{III.11}$$

and $a^2 = d^2 - b^2$. Here, $\mathrm{tr}(A)$ denotes the trace of matrix A.

Chen et al. [1] propose a refinement of the Ledoit-Wolf estimator, called the Rao-Blackwell Ledoit-Wolf estimator, which we denote by $\widehat{\mathbf{\Sigma}}_{RBLW}$, that has better asymptotical properties (as $n\to\infty$) than $\widehat{\mathbf{\Sigma}}_{LW}$. This is estimator is given by

$$\widehat{\mathbf{\Sigma}}_{RBLW} = \rho_{RBLW}\cdot\mathbf{I} + (1-\rho_{RBLW})\cdot\mathbf{S}, \tag{III.12}$$

where

$$\rho_{RBLW} = \min\left(\frac{\frac{n-2}{n}\cdot\mathrm{tr}\left(\mathbf{S}^2\right) + \mathrm{tr}^2\left(\mathbf{S}\right)}{(n+2)\cdot\left[\mathrm{tr}\left(\mathbf{S}^2\right) - \frac{\mathrm{tr}^2(\mathbf{S})}{p}\right]}, 1\right). \tag{III.13}$$

A second estimator proposed in [1] is the Oracle Approximating Shrinkage (OAS) estimator, which rely on a iterative procedure to provably converge to

$$\widehat{\mathbf{\Sigma}}_{OAS} = \rho_{OAS}\cdot\mathbf{I} + \left(1-\rho_{OAS)}\right)\cdot\mathbf{S}, \text{ with} \tag{III.14}$$

$$\rho_{OAS} = \min\left(\frac{(1-2/p)\cdot\mathrm{tr}\left(\mathbf{S}^2\right) + \mathrm{tr}^2\left(\mathbf{S}\right)}{(n+1-2/p)\left[\mathrm{tr}\left(\mathbf{S}^2\right) - \mathrm{tr}^2\left(\mathbf{S}\right)/p\right]}, 1\right). \tag{III.15}$$

The performance of the three estimators we presented is highly dependent on the data set. For example, in [1] $\widehat{\mathbf{\Sigma}}_{OAS}$ performs better than $\widehat{\mathbf{\Sigma}}_{RBLW}$ and $\widehat{\mathbf{\Sigma}}_{LW}$ for data coming from a fractional Brownian motion, but all three estimators perform comparably for Gaussian AR(1) processes (for small values of n).

We demonstrate the performance of the three estimators using the light-weight quasi-geostrophic (QG) model, which is representative of a realistic atmospheric or oceanic data assimilating system [12]. We consider the QG model on the 2-D cartesian domain $\Omega = [0,\,1]\times[0,\,1]$ estimate the covariance between 4×4 grid at 40 consecutive times (p =160). The QC model describes the motion of a fluid and is mathematically expressed as

$$q_t = -\psi_x - \epsilon\cdot\mathcal{J}\left(\psi, q\right) - A\cdot\Delta^3\psi + 2\cdot\pi\cdot\sin\left(2\cdot\pi\cdot y\right), \tag{III.16}$$

where q is the potential vorticity, ψ is the stream function, Δ is the Laplacian operator, $\mathcal{J}\left(\psi, q\right) = \psi_x\cdot q_y - \psi_y\cdot q_x$ is the Jacobian and x and y are the horizontal and vertical components, respectively. The coefficients A and ϵ are set to 2×10^{-12} and $A = 10^{-5}$, respectively.

The model uses homogeneous boundary conditions $\partial\Omega = \partial\psi = \partial q = 0$ and initial conditions of the form

$$\begin{aligned}\psi_{i,j} = {} & \sin(\alpha + \beta\cdot 4\cdot y_i\cdot x_j) + \cos(\alpha + \beta\cdot 2\cdot y_i\cdot x_j) \\ & + \sin(\alpha + \beta\cdot 2\cdot y_i\cdot x_j)\cdot\cos(\alpha + \beta\cdot 4\cdot y_i\cdot x_j),\end{aligned} \tag{III.17}$$

where (x_i, y_j), $1 \leq i, j \leq 64$, are the discretization points. Parameters α and β describe the shift and amplitude waves, respectively.

The samples are built by picking 4 points from the discretization of Ω at 40 consecutive times and randomly perturbing the initial conditions using $\alpha = 1 + |\mu_1|$ and $\beta = \mu_2$, where μ_1 and μ_2 are uniformly distributed random variables, $\mu_1 \sim \mathcal{U}(0, 10^{-4})$ and $\mu_2 \sim \mathcal{U}(0, 10^{-2})$. Each initial condition is propagated in time providing a sample associated with the 160 points of interest.

The quality of the estimators is infered based on the percentage relative improvement in average loss norm (PRIAL) that describes how much an estimator $\widehat{\mathbf{\Sigma}}$ improves the estimation of $\mathbf{\Sigma}$ with respect to $\mathbf{S}$ and is defined as

$$\delta\left(\mathbf{\Sigma}, \mathbf{S}, \widehat{\mathbf{\Sigma}}\right) = \frac{\mathbb{E}\left[\|\mathbf{\Sigma} - \mathbf{S}\|^2\right] - \mathbb{E}\left[\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\|^2\right]}{\mathbb{E}\left[\|\mathbf{\Sigma} - \mathbf{S}\|^2\right]}. \quad \text{(III.18)}$$

The true covariance matrix $\mathbf{\Sigma}$ is evaluated using 200 samples.

We evaluate the estimators using $n = 25$ and $n = 40$ samples and the PRIAL norms are shown in Table III-C. We note that $\widehat{\mathbf{\Sigma}}_{RBLW}$ offers the best improvement of the estimation over the sample covariance matrix and is more robust for different number of samples. In Figure 6 we show the structure of $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}_{RBLW}$ ($n = 25$ samples) and note that RBLW estimator preserves the structure of the true covariance $\mathbf{\Sigma}$.

TABLE I
PRIAL NORMS FOR THE ESTIMATORS FOR THE QG MODEL USING $n = 25$ AND $n = 40$ SAMPLES.

n	$\delta\left(\mathbf{\Sigma}, \mathbf{S}, \widehat{\mathbf{\Sigma}}_{LW}\right)$	$\delta\left(\mathbf{\Sigma}, \mathbf{S}, \widehat{\mathbf{\Sigma}}_{RBLW}\right)$	$\delta\left(\mathbf{\Sigma}, \mathbf{S}, \widehat{\mathbf{\Sigma}}_{OAS}\right)$
25	26%	66%	62%
40	35%	64%	61%

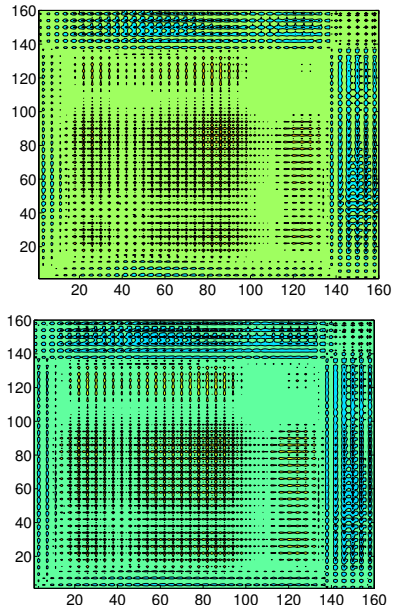

Fig. 6. True covariance matrix (top) and Rao-Blackwell Ledoit and Wolf estimator (bottom) for the sample size $n = 25$.

D. Integrative Numerical Study

To perform our benchmarks, we consider two strategies:

- **(Corr) :** This strategy computes the forward dispatch solution using scenarios that capture correlation. This strategy assumes that the ISO has correct spatio-temporal information.
- **(Indep) :** This strategy computes the forward dispatch solution using scenarios that do not capture correlation. We recall that the use of independent scenarios corresponds to the case in which each player submits to the ISO a set of scenarios created based on its own forecast, leaving the ISO without information about the correlation between players.

For both cases, we define as the **predicted cost** as the dispatch cost obtained from the solution of the corresponding dispatch problem. We then fix the forward decisions to evaluate cost at a new set of scenarios obtained by sampling the the distribution capturing correlation. We denote this cost as the **realized cost**.

We devised the simulations to also reveal the effect of dispatch cost and the number of scenarios used. For this we have solved the dispatch model with $S = \{4, 8, 16, 32, 64, 128, 256\}$ scenarios. Because the dispatch is a random variable, we computed error bands shown in Figure 7. For this we generated 256 batches of S scenarios for each $S = \{4, 8, 16, 32, 64, 128, 256\}$, solved the ED model for each batch and computed the mean and the standard deviation of the social surplus.

The sizes of the resulting optimization problems range from $14,635$ decision variables and $12,884$ constraints for $S = 4$ to $763,579$ decision variables and $704,372$ constraints for $S = 256$. The stochastic dispatch problems have a well-known "dual-block angular" structure that can be exploited to enable fast solutions in parallel using our solver PIPS-IPM [5], [6], [7], [8]. To compute the mean and standard deviation of the social surplus for each $S = \{4, 8, 16, 32, 64, 128, 256\}$ we solve 256 instances in parallel, each instance using S parallel processes. In these experiments we used "Intrepid" IBM BG/P and "Mira" IBM BG/Q supercomputers of Argonne National Laboratory. The Intrepid supercomputer has 40 racks with a total 40,960 nodes and a 3D torus high-performance interconnect; each BG/P node has a 850 MHz quad-core PowerPC processor and 2 GB of RAM. Mira is the replacement for Intrepid and consists of 48 racks, each of 1024 nodes, and a 5D torus interconnect. Each BG/Q node has 16 PowerPC A2 cores operating at 1600 Mhz and 16 GB of memory. In our simulations we have used up to $16,384$ nodes on each system (for the largest run corresponding to S=256). On Mira, the total execution times for solving 256 batches in parallel are in between 6 minutes (for $S = 4$) and 8 minutes (for $S = 256$). The slight increase in the execution times with S are primarily caused by I/O overhead, the optimization solution times remaining relatively constant (a bit more than 5 minutes for $S = 4$ and almost 6 minutes for $S = 256$).

From Figure 7 we can make the following observations:

- The (Indep) strategy yields a significant gap between the predicted and the realized costs. Interestingly, the gap is as large as $20,000$ USD/hr. The gap result because the (Indep) strategy fails at capturing the reality of atmospheric conditions at the moment of dispatching.
- The (Corr) strategy correctly predicts the dispatch cost. This can be observed by comparing the means.
- There is a positive social surplus gap between the (Corr) and the (Indep) strategies of about $10,967$ USD/hr or

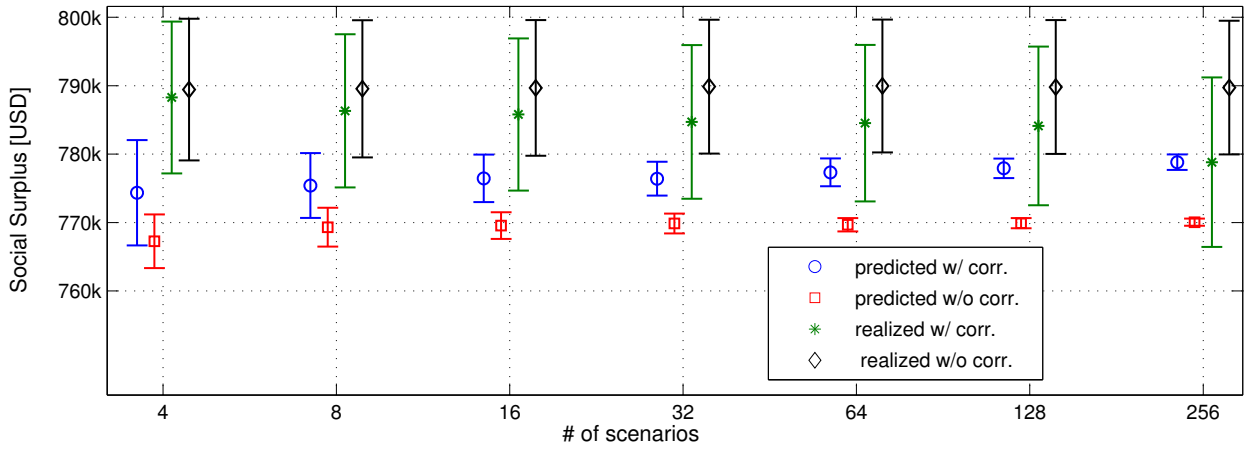

Fig. 7. Mean and 95% confidence intervals for dispatch cost predicted with correlated and independent scenarios. We also show "realized" dispatch costs, corresponding to costs obtained by implementing the predicted dispatch decisions under the correlated scenarios sampling scheme.

1.42%. This can add up to 100 million USD/yr. Furthermore, the gap does not close as the number of scenario increases, being consistent with the analytical dispatch model of Section II.

- The (Indep) and the (Corr) strategies exhibit different error bands and the bands at significantly different rates. For the (Corr) strategy, the error bands are small for 64 or more scenarios. In particular, the standard deviation of the dispatch cost is 0.45% and 0.36% when using 128 and 256 scenarios, respectively. This suggests that $\mathcal{O}(10^2)$ scenarios offer a good approximate social surplus even for a large number of wind farms. The (Indep) strategy underestimates the amount of scenarios needed.

IV. Conclusions and Future Work

We have demonstrated that neglecting correlation between multiple wind supply points can result in strong biases of dispatch cost. Our conclusions are drawn from a detailed study that incorporates high-resolution wind speed forecasts from numerical weather prediction models, covariance estimation techniques, and stochastic dispatch models.

Appendix

Lemma 1: $\mathbb{E}[X|X \le d] = -\sigma^2\phi(d) + \mu\Phi(d)$.

Proof: By direct computation we have

$$\begin{aligned}
&\mathbb{E}[X|X \le d] \\
&= \int_{-\infty}^{d} x\phi(x)dx = \\
&= -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{d} \left(-\frac{x-\mu}{\sigma^2} - \frac{\mu}{\sigma^2}\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= -\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\Big|_{\infty}^{d} \\
&\quad + \frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^{d} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= -\sigma^2\phi(d) + \mu\Phi(d).
\end{aligned}$$

We first show that $c_d(\sigma)$ given by (II.7) is strictly increasing. For this let $p(x,\sigma) = -\frac{(x-\mu)^2}{2\sigma^2}$, therefore $\phi(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp(p(x,\sigma))$. Observe that

$$\frac{d}{dx}p(x,\sigma) = -\frac{(x-\mu)}{\sigma^2}, \tag{A.19a}$$

$$\frac{d}{d\sigma}p(x,\sigma) = +\frac{(x-\mu)^2}{\sigma^3} \tag{A.19b}$$

$$\frac{d}{dx}\phi(x,\sigma) = \phi(x,\sigma)\cdot\frac{d}{dx}p(x,\sigma) \tag{A.19c}$$

$$\begin{aligned}
\frac{d}{d\sigma}\phi(x,\sigma) &= \phi(x,\sigma)\cdot\frac{d}{d\sigma}p(x,\sigma) - \frac{1}{\sigma^2\sqrt{2\pi}}\exp(p(x,\sigma)) \\
&= \frac{(x-\mu)^2}{\sigma^3}\cdot\phi(x,\sigma) - \frac{1}{\sigma}\phi(x,\sigma)
\end{aligned} \tag{A.19d}$$

The derivative of Φ with respect to σ can be computed as follows:

$$\begin{aligned}
&\frac{d}{d\sigma}\Phi(d,\sigma) \\
&= \frac{d}{d\sigma}\int_{-\infty}^{d}\phi(p(x,\sigma))dx \\
&= \int_{-\infty}^{d}\frac{d}{d\sigma}\phi(p(x,\sigma))dx \\
&\overset{\text{(A.19d)}}{=} \int_{-\infty}^{d}\frac{(x-\mu)^2}{\sigma^3}\cdot\phi(p(x,\sigma))dx - \frac{1}{\sigma}\int_{-\infty}^{d}\phi(p(x,\sigma))dx \\
&= -\int_{-\infty}^{d}\phi(p(x,\sigma))\cdot\left(-\frac{x-\mu}{\sigma^2}\right)\cdot\frac{x-\mu}{\sigma}dx - \frac{1}{\sigma}\Phi(d,\sigma) \\
&\overset{\text{(A.19a)}}{=} -\int_{-\infty}^{d}\phi(p(x,\sigma))\cdot\frac{d}{dx}p(x,\sigma)\cdot\frac{x-\mu}{\sigma}dx - \frac{1}{\sigma}\Phi(d,\sigma) \\
&\overset{\text{(A.19c)}}{=} -\int_{-\infty}^{d}\frac{d}{dx}\phi(p(x,\sigma))\cdot\frac{x-\mu}{\sigma}dx - \frac{1}{\sigma}\Phi(d,\sigma) \\
&= -\phi(p(x,\sigma))\cdot\frac{x-\mu}{\sigma}\Big|_{-\infty}^{d} \\
&\quad + \frac{1}{\sigma}\int_{-\infty}^{d}\phi(p(x,\sigma))dx - \frac{1}{\sigma}\Phi(d,\sigma) \\
&= -\frac{d-\mu}{\sigma}\cdot\phi(p(d,\sigma)).
\end{aligned} \tag{A.20}$$

Finally, by differentiating (II.7) and using (A.19d) and (A.20), compute

$$\begin{aligned}\frac{d}{d\sigma}c_d(\sigma) &= (p_{th}-p_w)\left((d-\mu)\frac{d}{d\sigma}\Phi(d,\sigma)+2\sigma\phi(x,d)\right.\\ &\qquad\left.+\sigma^2\frac{d}{d\sigma}\phi(d,\sigma)\right)\\ &= (p_{th}-p_w)\left(-\frac{(d-\mu)^2}{\sigma}\cdot\phi(d,\sigma)+2\sigma\phi(d,\sigma)\right.\\ &\qquad\left.+\frac{(d-\mu)^2}{\sigma}\cdot\phi(d,\sigma)-\sigma\phi(d,\sigma)\right)\\ &= \sigma(p_{th}-p_w)\phi(d,\sigma)>0, \qquad (A.21)\end{aligned}$$

which shows that $c_d(\sigma)$ is strictly increasing. Since $\sigma(\rho)=\sqrt{\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2}$ is also a strictly increasing function of ρ, we conclude that c_d is also strictly increasing function of the correlation coefficient ρ.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, under Contract No. DE-AC02-06CH11357.

REFERENCES

[1] YILUN CHEN, A. WIESEL, AND A.O. HERO, *Shrinkage estimation of high dimensional covariance matrices*, in Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, 2009, pp. 2937–2940.

[2] E.M. CONSTANTINESCU, V.M. ZAVALA, M. ROCKLIN, SANGMIN LEE, AND M. ANITESCU, *A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation*, IEEE Transactions on Power Systems, 26 (2011), pp. 431–441.

[3] EMIL M CONSTANTINESCU, VICTOR M ZAVALA, MATTHEW ROCKLIN, SANGMIN LEE, AND MIHAI ANITESCU, *A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation*, Power Systems, IEEE Transactions on, 26 (2011), pp. 431–441.

[4] OLIVIER LEDOIT AND MICHAEL WOLF, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis, 88 (2004), pp. 365 – 411.

[5] MILES LUBIN, COSMIN G. PETRA, AND MIHAI ANITESCU, *The parallel solution of dense saddle-point linear systems arising in stochastic programming*, Optimization Methods and Software, 27 (2012), pp. 845–864.

[6] MILES LUBIN, COSMIN G. PETRA, MIHAI ANITESCU, AND VICTOR ZAVALA, *Scalable stochastic optimization of complex energy systems*, in Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC '11, New York, USA, 2011, ACM, pp. 64:1–64:64.

[7] COSMIN G. PETRA AND MIHAI ANITESCU, *A preconditioning technique for Schur complement systems arising in stochastic optimization*, Computational Optimization and Applications, 52 (2012), pp. 315–344.

[8] COSMIN G. PETRA, OLAF SCHENK, MILES LUBIN, AND KLAUS GÄRTNER, *An augmented incomplete factorization approach for computing the Schur complement in stochastic optimization*, accepted to SIAM Journal on Scientific Computing, (2014).

[9] P. PINSON, H. MADSEN, H. A. NIELSEN, G. PAPAEFTHYMIOU, AND B. KLCKL, *From probabilistic forecasts to statistical scenarios of short-term wind power production*, Wind Energy, 12 (2009), pp. 51–62.

[10] GEOFFREY PRITCHARD, GOLBON ZAKERI, AND ANDREW PHILPOTT, *A single-settlement, energy-only electric power market for unpredictable and intermittent participants*, Operations Research, 58 (2010), pp. 1210–1219.

[11] ———, *A single-settlement, energy-only electric power market for unpredictable and intermittent participants*, Operations Research, 58 (2010), pp. 1210–1219.

[12] PAVEL SAKOV AND PETER R. OKE, *A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters*, Tellus A, 60 (2008), pp. 361–371.

[13] W.C. SKAMAROCK, J.B. KLEMP, J. DUDHIA, D.O. GILL, D.M. BARKER, M.G. DUDA, X.-Y. HUANG, W. WANG, AND J.G. POWERS, *A description of the Advanced Research WRF version 3*, Tech. Report Tech Notes-475+ STR, NCAR, 2008.

[14] J. SUMAILI, H. KEKO, V. MIRANDA, ZHI ZHOU, A. BOTTERUD, AND JIANHUI WANG, *Finding representative wind power scenarios and their probabilities for stochastic models*, in 16th International Conference on Intelligent System Application to Power Systems, Sept 2011, pp. 1–6.

[15] J. TASTU, P. PINSON, AND H. MADSEN, *Space-time trajectories of wind power generation: Parameterized precision matrices under a gaussian copula approach*, in Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension, in press, 2014.

[16] M. WYTOCK AND J.Z. KOLTER, *Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields*, in 2013 IEEE 52nd Annual Conference on Decision and Control, Dec 2013, pp. 1019–1024.

[17] V. M. ZAVALA, J. BIRGE, AND M. ANITESCU, *A stochastic market clearing formulation with consistent pricing properties*, Under Review (2014).